nature physics

ARTICLES

https://doi.org/10.1038/s41567-020-0990-x

A strong no-go theorem on the Wigner's friend paradox

Kok-Wei Bong[1,4], Aníbal Utreras-Alarcón[1,4], Farzad Ghafari[1], Yeong-Cherng Liang[2], Nora Tischler[1]✉, Eric G. Cavalcanti[3]✉, Geoff J. Pryde[1] and Howard M. Wiseman[1]

Does quantum theory apply at all scales, including that of observers? New light on this fundamental question has recently been shed through a resurgence of interest in the long-standing Wigner's friend paradox. This is a thought experiment addressing the quantum measurement problem—the difficulty of reconciling the (unitary, deterministic) evolution of isolated systems and the (non-unitary, probabilistic) state update after a measurement. Here, by building on a scenario with two separated but entangled friends introduced by Brukner, we prove that if quantum evolution is controllable on the scale of an observer, then one of 'No-Superdeterminism', 'Locality' or 'Absoluteness of Observed Events'—that every observed event exists absolutely, not relatively—must be false. We show that although the violation of Bell-type inequalities in such scenarios is not in general sufficient to demonstrate the contradiction between those three assumptions, new inequalities can be derived, in a theory-independent manner, that are violated by quantum correlations. This is demonstrated in a proof-of-principle experiment where a photon's path is deemed an observer. We discuss how this new theorem places strictly stronger constraints on physical reality than Bell's theorem.

Wigner's friend[1] is a thought experiment that illustrates what is perhaps the thorniest foundational problem in quantum theory: the measurement problem[2,3]. In the thought experiment, we consider an observer (the 'friend') who performs a measurement on a quantum system. In accordance with the state update rule, the friend assigns the eigenstate corresponding to their observed outcome to the measured system. The friend is assumed to be inside an isolated laboratory that can be coherently controlled by a second experimenter, Wigner, who is capable of performing arbitrary quantum operations on the friend's laboratory and all of its contents. Although this may be possible, in principle, it would be a truly Herculean task if the friend were a macroscopic observer like a human, as we have chosen for our illustrations and discussions below. For this reason, Wigner is often called a 'superobserver'. However, there is good reason to think that quantum mechanics would allow control of the type required if the friend were an artificial intelligence algorithm in a simulated environment running in a large quantum computer. Wigner describes the laboratory and all of its contents as a unitarily evolving quantum state, in accordance with the rule for state evolution applicable to isolated systems. The case when the friend's system is prepared in a superposition state leads to an apparent contradiction between the friend's perspective and that of Wigner, who does not ascribe a well-defined value to the outcome associated with his friend's observation. For a more in-depth description of the Wigner's friend thought experiment, see Supplementary Section A.

Although decoherence can 'save the appearances' by explaining the suppression of quantum effects at the macroscopic level, it cannot solve the measurement problem: 'we are still left with a multitude of (albeit individually well-localized quasiclassical) components of the wave function, and we need to supplement or otherwise to interpret this situation in order to explain why and how single outcomes are perceived'[2]. Proposed resolutions have radical implications: they either reject the idea that measurement outcomes have single, observer-independent values[4–7] or postulate faster-than-light[8,9] or retrocausal effects[10,11] at a hidden variable level. Alternatively, some theories postulate mechanisms to avoid macroscopic superpositions, such as modifications to unitary quantum dynamics[12] or gravity-induced collapse[13]. Here we rigorously demonstrate that radical revisions of such types are in fact required.

Our work is inspired by the recent surge of renewed interest in the Wigner's friend problem[14–20]. In particular, Brukner[14] introduced an extended Wigner's friend scenario (EWFS) with two spatially separated laboratories, each containing a friend, accompanied by a superobserver who can perform various measurements on their friend's laboratory. Each friend measures half of an entangled pair of systems, establishing correlations between the results of the superobservers' subsequent measurements.

In the context of this EWFS, Brukner[14,15,20] considered three assumptions: 'freedom of choice', 'locality' (in the sense of 'parameter independence'[21]) and 'observer-independent facts' (OIFs). The last of these means that propositions about all observables that might be measured (by an observer or a superobserver) are 'assigned a truth value independently of which measurement Wigner performs'[14].

In other words, the OIF assumption is equivalent to the assumption of Kochen–Specker non-contextuality[22,23] (KSNC). From these assumptions, Brukner derived a Bell inequality for the correlations of the superobservers' results, which could be violated in quantum mechanics (if the superobservers could suitably manipulate the quantum state of the observers). A recent six-photon experiment[17], using a set-up where the role of each friend is played by a single

[1]Centre for Quantum Computation and Communication Technology (Australian Research Council), Centre for Quantum Dynamics, Griffith University, Brisbane, Queensland, Australia. [2]Department of Physics and Center for Quantum Frontiers of Research & Technology (QFort), National Cheng Kung University, Tainan, Taiwan. [3]Centre for Quantum Computation and Communication Technology (Australian Research Council), Centre for Quantum Dynamics, Griffith University, Southport, Queensland, Australia. [4]These authors contributed equally: Kok-Wei Bong, Aníbal Utreras-Alarcón. ✉e-mail: n.tischler@griffith.edu.au; e.cavalcanti@griffith.edu.au

photon, successfully violated such a Bell inequality derived from Brukner's assumptions.

Although the EWFS background for this result was novel, the derived Bell inequality can be obtained from the assumptions of 'freedom of choice' and KSNC, without considering the friends' observations, and without using 'locality' (which follows from Bell's stronger notion of local causality[24], which in turn follows from KSNC in any Bell scenario[25]). Furthermore, the Kochen–Specker theorem[22] already establishes that KSNC + 'freedom of choice' leads to contradictions with quantum theory. As discussed in refs. [19,20,26], this casts doubt on the implications of Brukner's theorem with regard to any assumption specifically about the objectivity of the friends' observations—one can respond to Brukner's theorem simply by maintaining that 'unperformed experiments have no results'[27].

Nevertheless, there is a subtle but important difference between a standard Bell scenario in which one of two incompatible observables are chosen at random to be measured by each party and the scenario introduced by Brukner. In the latter, in one of four experimental runs, all four observables involved in the experiment are being measured—one by each observer in the scenario. This suggests that the counterfactual reasoning in the OIF/KSNC assumption could be avoided by replacing it with a suitable weaker assumption. Indeed, Brukner discusses a weaker assumption—'that Wigner's and Wigner's friend's facts coexist'—before settling on 'The assumption of 'observer-independent facts' [which] is a stronger condition'[14].

In this Article we derive a new theorem, based on the intuition in the preceding paragraph around Brukner's EWFS. It uses metaphysical assumptions (that is, assumptions about physical theories) that are strictly weaker than those of Bell's theorem or Kochen–Specker contextuality theorems, and thus opens a new direction in experimental metaphysics. Our first two assumptions are, as per Brukner, 'freedom of choice' (which we make more formal using the concept of 'No-Superdeterminism' defined in ref. [24]) and 'Locality' (in the same sense as Brukner; see also ref. [24]). Our third assumption is 'Absoluteness of Observed Events' (AOE), which is that an observed event is a real single event and not relative to anything or anyone. Note that capitalization is used for assumptions formally defined in this paper.

Unlike OIF, AOE makes no claim about hypothetical measurements that were not actually performed in a given run. Furthermore, AOE is necessarily (though often implicitly) assumed even in standard Bell experiments[24]. For convenience, we will call the conjunction of these three assumptions 'Local Friendliness' (LF). This enables us to state our theorem.

***Theorem* 1**: If a superobserver can perform arbitrary quantum operations on an observer and its environment, then no physical theory can satisfy Local Friendliness.

By a 'physical theory' we mean any theory that correctly predicts the correlations between the outcomes observed by the superobservers Alice and Bob (Fig. 1), who can communicate after their experiments are performed and evaluate those correlations. The proof of Theorem 1 proceeds by showing that LF implies a set of constraints on those correlations (that we call 'LF inequalities') that can, in principle, be violated by quantum predictions for an EWFS scenario. Thus, like Bell's theorem and Brukner's theorem, our theorem is theory-independent—we use (like Bell and Brukner) quantum mechanics as a guide for what may be seen in experiments, but the metaphysical conclusions hold for any theory if those predictions are realized in the laboratory. (This is unlike the theorem of ref. [16], which is a statement about the standard theory of quantum mechanics.) Note also that, unlike in Brukner's theorem, all three assumptions going into LF are essential for the theorem, and so are the friends' observations.

For the specific EWFS Brukner considered—involving two binary-outcome measurement choices per superobserver—the set of correlations allowed by our LF assumption is identical to the set allowed by the assumptions of Bell's theorem, commonly referred to as the local hidden variable (LHV) correlations. However, in general, LF and LHV do not give identical constraints. Indeed, already for a slightly more complicated EWFS with three binary-outcome measurement choices per superobserver, we show that the set of LF correlations is a strict superset of the set of LHV correlations. Moreover, it is possible for quantum correlations to violate a Bell inequality (an inequality bounding the set of LHV correlations) while satisfying all of the LF inequalities. We also prove that the new LF inequalities we derive can nevertheless be violated by quantum correlations. We demonstrate these facts in an experimental simulation where the friends are represented by photon paths.

We now proceed to explain the EWFS in more detail, before presenting our results and discussing their implications.

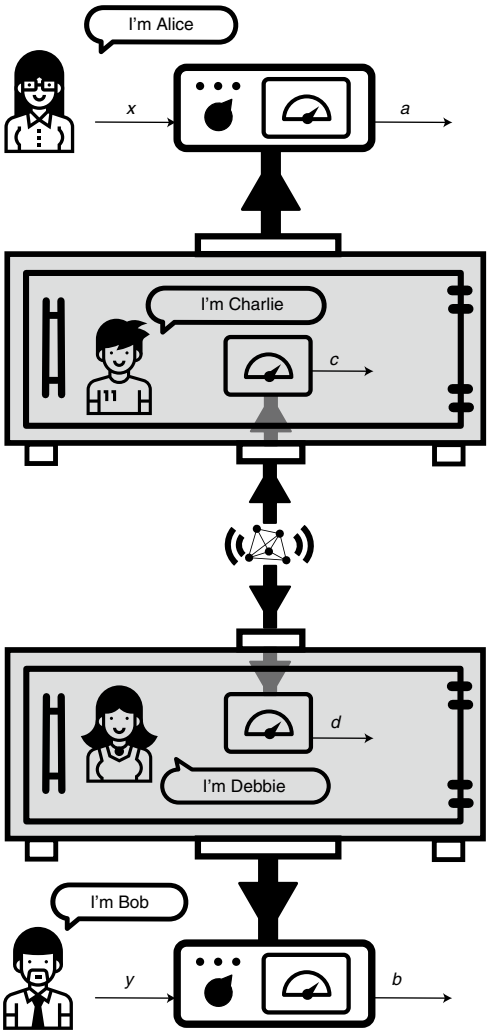

Fig. 1 | Concept of the extended Wigner's friend scenario. The friends, Charlie and Debbie, measure a pair of particles prepared in an entangled state, producing the outcomes labelled *c* and *d*, respectively (from their perspective). The superobservers, Alice and Bob, perform space-like separated measurements labelled *x* and *y*, with outcomes labelled *a* and *b*, on the entire contents of the laboratories containing Charlie and Debbie, respectively. Credit: Icons of people, Eucalyp Studio under a Creative Commons licence (https://creativecommons.org/licenses/by/3.0/).

The extended Wigner's friend scenario. Let us consider the bipartite version of the Wigner's friend experiment that was introduced

by Brukner, involving two superobservers, Alice and Bob, and their respective friends, Charlie and Debbie (Fig. 1). Charlie and Debbie each have one particle from an entangled pair, and make a measurement on it, yielding outcomes c and d, respectively.

In each iteration of the experiment, Alice and Bob randomly and independently choose one out of $N \geq 2$ measurements to be performed in space-like separated regions subsequent to a space-like hypersurface containing the measurements of both Charlie and Debbie, as shown in Fig. 2d. The settings are respectively labelled $x \in \{1, \ldots, N\}$ and $y \in \{1, \ldots, N\}$, with corresponding outcomes a and b (we do not assume anything about the number of possible outcomes at this stage). For the specific EWFS depicted in Fig. 2, if $x = 1$, Alice simply opens Charlie's laboratory and directly asks him for his outcome c, then assigns her own outcome as $a = c$, as shown in Fig. 2a.

For $x \in \{2, \ldots, N\}$, Alice performs a different measurement on Charlie's laboratory as a whole. In particular, we will consider measurements such that Alice restores the laboratory to a previous state, thereby erasing Charlie's memory (Fig. 2b), and then proceeds to measure the particle directly (Fig. 2c). Bob and Debbie operate in a similar fashion.

From this experiment we can measure (as frequencies) the empirical probabilities $\wp(ab|xy)$, using only the information available at the end of the experiment, namely, the values for a, b, x and y. Unless $x = 1$, all records for the value of c are erased when Alice performs her measurement, so in general that information cannot be accessed at the end of the experiment, and likewise with the value of d on Bob's side. A detailed quantum-mechanical description of the EWFS is provided in the Methods.

Formalization of the LF assumptions. Within a bipartite Wigner's friend experiment, what constraints do the LF assumptions imply for the probabilities $\wp(ab|xy)$ observed by Alice and Bob for outcomes a and b, given settings x and y? To determine this rigorously we need to formalize our three assumptions.

Assumption 1 (Absoluteness of Observed Events (AOE)): An observed event is a real single event, and not relative to anything or anyone.

In an EWFS, the assumption of AOE implies that, in each run of the experiment—that is, given that Alice has performed measurement x and Bob has performed measurement y on some pair of systems—there exists a well-defined value for the outcome observed by each observer, that is, for a, b, c and d. Formally, this implies that there exists a theoretical joint probability distribution $P(abcd|xy)$ from which the empirical probability $\wp(ab|xy)$ can be obtained while also ensuring that the observed outcomes for $x, y = 1$ are consistent between the superobservers and the friends.

- AOE (in the EWFS of Fig. 2): $\exists\, P(abcd|xy)$ s.t.

i. $\wp(ab|xy) = \Sigma_{c,d} P(abcd|xy)\ \forall\ a, b, x, y$
ii. $P(a|cd, x = 1, y) = \delta_{a,c}\ \forall\ a, c, d, y$
iii. $P(b|cd, x, y = 1) = \delta_{b,d}\ \forall\ b, c, d, x$

Here, we do not assume that all statements about results have truth values independently of which measurement 'Wigner' (whom we call Alice) performs. Instead, the assumption of AOE only entails assigning truth values to propositions about observed outcomes. In particular, Alice's measurement outcome A_x (which in our notation corresponds to the value of a when she performs the measurement labelled by x) for $x \neq 1$ has a value only when she performs that measurement. However, A_1 is different in that it has a value even when $x \neq 1$, because it is encoded in c, which is actually measured by Charlie in every run. All this is in keeping with Peres' dictum 'unperformed experiments have no results'[27]; AOE is the assumption that performed experiments have observer-independent (that is, absolute) results.

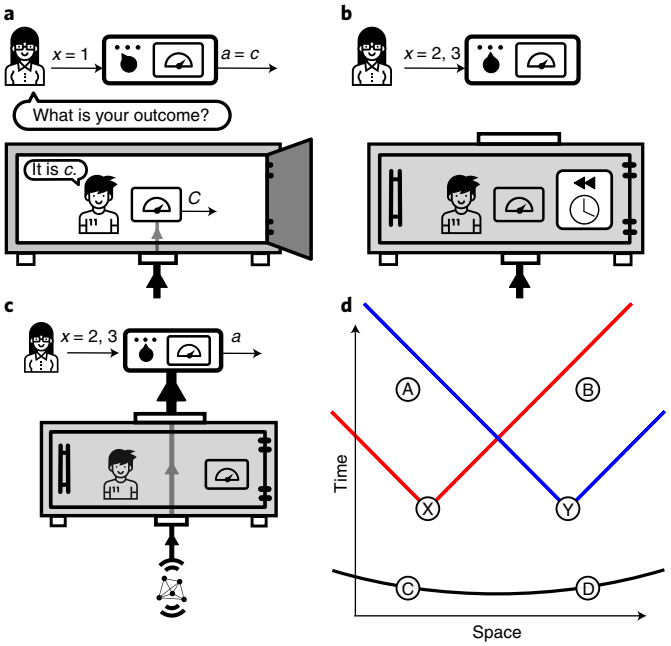

Fig. 2 | A specific bipartite Wigner's friend experiment. a, When $x = 1$, Alice opens Charlie's laboratory and asks him his outcome. **b**, Alternatively, for $x = 2, 3$, she may restore the laboratory to a previous state. **c**, She then proceeds to ignore Charlie, and performs a measurement directly on the particle. **d**, Space–time diagram illustrating the time ordering of the events within the experiment—C (D) is Charlie's (Debbie's) measurement, X (Y) is the event of Alice's (Bob's) choice of measurement setting, A (B) is Alice's (Bob's) measurement. Credit: Icons of people, Eucalyp Studio under a Creative Commons licence (https://creativecommons.org/licenses/by/3.0/).

The No-Superdeterminism assumption is a formalization of the assumption of 'freedom of choice' used in derivations of Bell inequalities. It is the assumption that the experimental settings can be chosen freely, that is, uncorrelated with any relevant variables prior to that choice. For added clarity, here we formulate it, following ref. [24] as follows.

Assumption 2 (No-Superdeterminism (NSD)): Any set of events on a space-like hypersurface is uncorrelated with any set of freely chosen actions subsequent to that space-like hypersurface.

In the EWFS, this implies that c and d are independent of the choices x and y:

- NSD (in the EWFS and under Assumption 1): $P(cd|xy) = P(cd)\ \forall\ c, d, x, y$

Finally, the assumption of Locality prohibits the influence of a local setting (such as x) on a distant outcome (such as b). It is the assumption that Bell, in 1964[28], and many others subsequently, also called 'locality'[24], and which Shimony called 'parameter independence'[21]; that is, in the formalization of ref. [24], the following assumption.

Assumption 3 (Locality (L)): The probability of an observable event e is unchanged by conditioning on a space-like-separated free choice z, even if it is already conditioned on other events not in the future light-cone of z.

In the EWFS, this implies:

- L (in the EWFS and under Assumption 1):

$P(a|cdxy) = P(a|cdx)\ \forall\ a, c, d, x, y$
$P(b|cdxy) = P(b|cdy)\ \forall\ b, c, d, x, y$

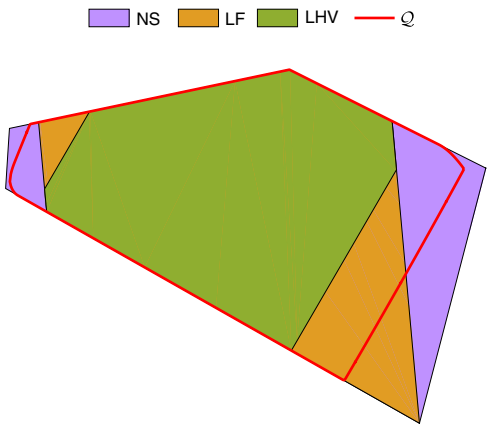

Fig. 3 | A two-dimensional slice of the space of correlations, illustrating the correlations discussed in this work. The solid areas depict a hierarchy of models: LHV[28] correlations (green) are a subset of LF correlations (green and orange), which in turn are a subset of no-signalling[39] correlations (NS, green, orange and purple). The red line bounds the correlations allowed by quantum theory on this slice. Note that, although the set of quantum correlations includes the LHV set, it does not include the LF set. Further details of this plot are discussed in Supplementary Section C).

Note that one could alternatively formulate Assumptions 2 and 3 as a single, equivalent assumption, which has previously been coined 'local agency' in the context of Bell's theorem[24]. Within the definitions of L and NSD, c, d play the formal role of the hidden variables λ in the usual derivation of Bell inequalities. However, we emphasize again that those correspond to observed events, and note that we make no assumption about hidden variables predetermining all measurement outcomes.

We call the set of correlations $\wp(ab|xy)$ that satisfy Assumptions 1–3 the LF correlations.

Properties of LF correlations. Our key findings about the properties of LF correlations are as follows. (1) LF correlations are a superset of LHV correlations, and in general a strict superset, as we will show quantitatively in the next section. (2) LF correlations can always be characterized by a finite set of inequalities. (3) For $N=2$ measurement settings and any number of measurement outcomes, LF correlations are the same as LHV correlations. (4) For $N=3$ measurement settings and $O=2$ outcomes, we fully characterize the LF correlations by deriving the associated inequalities and we show that they are a strict superset of LHV correlations (as illustrated in Fig. 3). We provide the derivations to these results in the Methods.

For $N=3$ measurement settings and $O=2$ outcomes, the set of LF correlations is a polytope with 932 facets. The facets can be grouped into nine inequivalent classes, each represented by a different inequality (provided in the Methods). These classes can be further grouped into categories, according to the measurement settings involved, and whether the facets are Bell facets[29]. In Table 1, we list the categories of LF facets, ignoring all positivity facets, that is, the constraints that probabilities cannot be negative.

Quantum violations. We now search for quantum violations of the LF inequalities. To demonstrate that the set of LF correlations is strictly larger than the LHV correlations, we seek a state and measurement choices such that a violation of a Bell non-LF inequality is exhibited without a violation in any of the LF inequalities. For experimental convenience, we consider two-qubit photon polarization states of the form

$$\rho_\mu = \mu|\Phi^-\rangle\langle\Phi^-| + \frac{1-\mu}{2}(|HV\rangle\langle HV| + |VH\rangle\langle VH|) \qquad (1)$$

Table 1 | Categories of inequalities for three binary-outcome measurement settings per party

Label	Measurement settings	LF inequality?	Bell facet?
Brukner	(1 i, 1 j)	Yes	Yes
Semi-Brukner	(1 i, 2 3)	Yes	Yes
Bell non-LF	(2 3, 2 3)	No	Yes
I_{3322}	(1 2 3, 1 2 3)	Yes	Yes
Genuine LF	(1 2 3, 1 2 3)	Yes	No

The column 'Measurement settings' refers to the settings that appear in each inequality, with i, $j \in \{2, 3\}$. The third column specifies whether it is an LF inequality, and the fourth column specifies whether it is a facet of the Bell polytope. Each category represents inequalities with the same form up to arbitrary relabelling of measurement settings (for $i, j \neq 1$), outcomes and parties. The labels referring to each inequality are 'Genuine LF' for inequalities that are not facets of the LHV polytope for this scenario, 'I_{3322}' for a type of Bell facet for the case of three binary-outcome measurement settings per party[37] and 'Brukner', 'Semi-Brukner' and 'Bell non-LF' are inequivalent classes of CHSH-type inequalities[38]. Brukner inequalities are the type of inequalities considered by Brukner[14]. A semi-Brukner inequality has a simpler experimental realization than a Brukner inequality, as it only requires one of the parties to measure a friend (setting 1). Bell non-LF inequalities are Bell facets, but unlike the other categories, are not facets of LF.

where $|\Phi^-\rangle = (|HV\rangle - |VH\rangle)/\sqrt{2}$, $0 \leq \mu \leq 1$, and H and V denote horizontal and vertical polarizations, respectively.

In Fig. 4 we display quantum violations for inequalities of all the categories in Table 1 for states ρ_μ. The specific inequalities and measurements considered are described in the Methods. Each of the inequalities considered is violated by some ρ_μ. In addition, we determine the strongest violations of the genuine LF inequalities allowed in quantum theory; those results are provided in Supplementary Section B.

In summary, if quantum measurements can be coherently performed at the level of observers, quantum mechanics predicts the violation of the LF inequalities in EWFSs. This proves Theorem 1.

Experiment. We study the EWFS with three measurement settings ($N=3$) in an experiment where the systems distributed between the two laboratories are polarization-encoded photons, the friends are photon paths within the set-up and the measurements by the superobservers are photon-detection measurements. Because the qubit composed of the two photon paths that represents each of our friends would not typically be considered a macroscopic, sentient observer as originally envisioned by Wigner, our experiment is best described as a proof-of-principle version of the EWFS. The experiment lets us demonstrate the key properties of LF inequalities and its results generalize provided that quantum evolution is, in principle, controllable on the scale of an observer. A fully rigorous demonstration that the LF assumptions are untenable would require, in addition to a more plausible 'observer', implementing shot-by-shot randomized measurement settings and closing separation, efficiency and freedom-of-choice loopholes, similarly to the case of Bell inequality violations[30–32].

Our experimental set-up, which comprises a photon source and a measurement section, is illustrated in Fig. 5. The photon source, shown in the left half of Fig. 5, is designed to generate the quantum state ρ_μ of equation (1) with a tunable μ parameter. Details about this spontaneous parametric downconversion source are provided in the Methods.

The measurement section of the experimental set-up, shown in the right half of Fig. 5, consists of two copies of an apparatus, one belonging to Alice and Charlie and the other to Bob and Debbie. The measurement section serves two purposes. The first is to perform quantum state tomography to characterize the generated quantum state, as detailed in the Methods and Supplementary Section D. The second purpose is to perform the measurements of

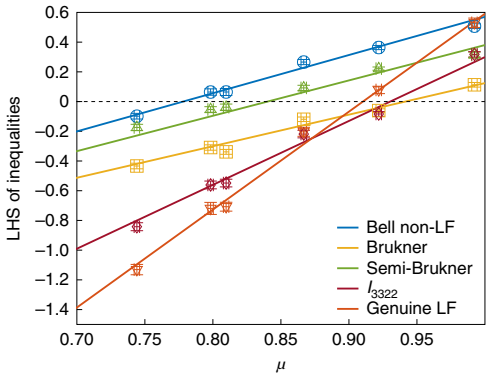

Fig. 4 | Results for the left-hand sides of Bell and LF inequalities for different quantum states. The parameter μ is the pure-state fraction of the quantum state in equation (1). The measurements and inequalities considered are provided in the Methods, using the labels introduced in Table 1. The dashed line in the plot represents the bound above which a violation occurs. The solid lines are theory predictions and the symbols represent experimental data. The uncertainties for the data points represent ±1 standard deviations, calculated from Monte Carlo simulations using 100 samples of Poisson-distributed photon counts. Figure reproduced with permission from ref. [40], SPIE.

the four observers, Charlie, Debbie, Alice and Bob. The friend's projective polarization measurement result is encoded in the photon path after the QWP, HWP and beam displacer BD1. Alice and Bob can perform different positive-operator valued measures (POVMs) on their respective system + friend, which depend on their measurement settings and are described in the Methods.

The experimental results are shown in Fig. 4. The μ values cover the full range of interest, from none of the inequalities being violated (at low μ), to the violation of all inequalities (at high μ). The experimental data demonstrate the sequential violations of the Bell non-LF, semi-Brukner and genuine LF inequalities. The data points corresponding to $\mu = 0.80$ and $\mu = 0.81$ are of particular significance, as they demonstrate that it is possible to violate Bell inequalities without violating any LF inequalities. (We can be confident of this because we verified that none of the 932 LF inequalities is violated.)

This means that the correlations consistent with LF assumptions are a superset of the correlations consistent with an LHV model. The case of $\mu = 0.87$ is the first of the plotted datasets where a contradiction with the LF assumptions occurs, through the first violation of an inequality associated with LF. Finally, the two highest μ values verify that the genuine LF inequality can also be violated. All the experimental data points, except for the case of $\mu = 0.81$, are at least two standard deviations away from 0, thus attesting the violation or non-violation of the inequalities with statistical significance. This covers all the regions we show in terms of (non-)violation of different inequalities, because the dataset at $\mu = 0.81$ belongs to the same region as $\mu = 0.80$. Along with the experimental data, the results predicted for the design measurement directions and input states of equation (1) are shown by solid lines. However, because the inequalities are device-independent, our conclusions are independent of which states and measurement directions were actually employed in the experiment.

Implications of violating LF inequalities. It is interesting to compare the assumptions that go into the LF no-go theorem with those for Bell's theorem. First, we note that the AOE assumption is implicit in the derivation of Bell inequalities (see ref. [24] for a derivation in which it is explicitly included as 'macroreality'). If, as is common, we also formulate Bell's theorem using the other two assumptions of LF, namely NSD and L, then an additional assumption is required. The minimal extra assumption required is 'outcome independence'[21], which in the bipartite scenario is the requirement that $P(a|bxy\lambda) = P(a|xy\lambda)$, $P(b|axy\lambda) = P(b|xy\lambda) \forall\, a, b, x, y, \lambda$ (c.f. the definition of L in the section 'Formalization of the LF assumptions'). Hence, the LF assumption is strictly weaker than the set of assumptions for Bell inequalities. Thus, the conclusions we could derive from an empirical violation of the LF inequalities are strictly stronger.

One popular way to accommodate the violation of Bell inequalities is to reject outcome independence (which is violated by operational quantum theory[24]) while maintaining L and NSD. Our theorem shows that this strategy does not extend to the EWFS. If the LF inequalities were violated empirically, then, to maintain L and NSD, one would have to reject AOE.

It is important to keep in mind that it is much harder to satisfy the conditions for an experimental violation of the LF inequalities than of Bell inequalities. A fully convincing demonstration would require a strong justification for the attribution of a 'fact' to the friend's measurement. This, of course, depends on what counts as an 'observer' (and as a 'measurement'). Because conducting this kind of experiment with human beings is physically impractical, what do we learn from experiments with simpler 'friends'?

Wigner's own conclusion from his thought experiment was that the collapse of the wave function should happen at least before it reaches the level of an 'observer'. The concept of an 'observer', however, is a fuzzy one. Objective collapse theories[12] attempt to restore the absolute reality of observed events by postulating modifications to the quantum dynamics to guarantee that collapse occurs before a quantum superposition reaches the macroscopic level. In other words, this resolution requires observed events to correspond to sufficiently macroscopic irreversible physical processes. In that case, the LF inequalities would not be violated with actual observers. Clearly, our experiment (and that of ref. [17]) did not probe collapse theories. Therefore, an open possibility is that the LF assumptions are valid, but that nature forever forbids the observation of violation of LF inequalities with observers, whether because of objective collapse or some other limitation on coherent quantum control.

A challenge to the above resolution of the LF no-go theorem could come from experiments involving AI (artificial intelligence) agents in a quantum computer. If universal quantum computation and strong AI are both physically possible, it should be possible to realize quantum coherent simulations of an observer and its (virtual) environment, and realize an extended Wigner's friend experiment. The experiment can even be conducted with a single friend, which would already allow testing semi-Brukner inequalities (equation (18)). Towards the goal of challenging the LF no-go theorem, experiments can test agents of increasing complexity; an experimental violation of LF inequalities with a given class of physical systems as 'friends' implies that either the LF assumptions are false or that class of friends is not an 'observer'.

Among interpretations of quantum mechanics that allow, in principle, the violation of LF inequalities, Theorem 1 can be accommodated in different ways. Interpretations that reject AOE include QBism[6,7], the relational interpretation[5] and the many-worlds interpretation[4]. Bohmian mechanics[8,9] violates L but not the other assumptions. There are some advocates for giving up NSD (either due to retrocausality[10], superdeterminism[11] or other mechanisms), but, as yet, no such theory has been proposed that reproduces all the predictions of quantum mechanics.

Finally, it was brought to our attention that the LF polytopes have been independently studied under the name of 'partially deterministic polytopes'[33], from an information-theoretic motivation: they are connected to the problem of device-independent randomness certification (see, for example, refs. [34–36] and references therein) in the presence of no-signalling adversaries.

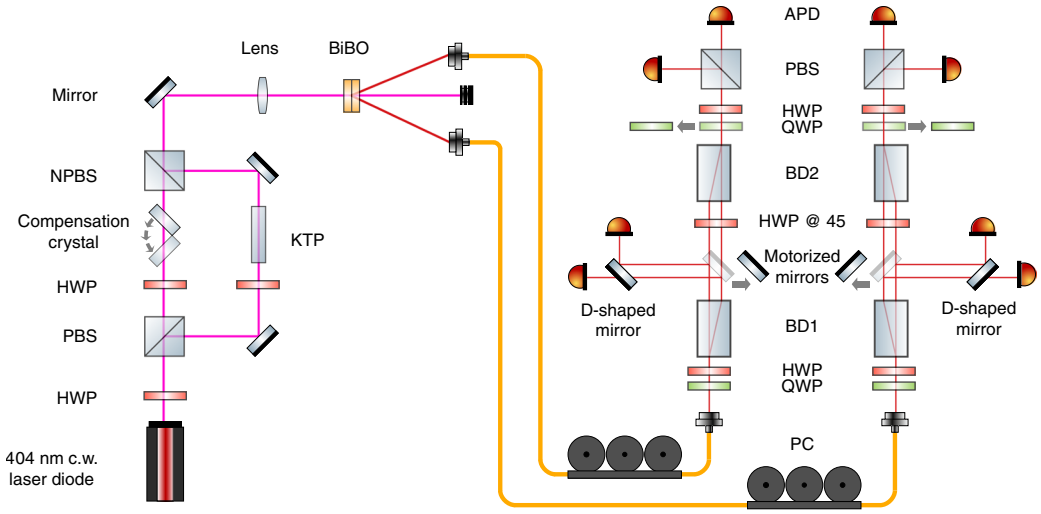

Fig. 5 | Experimental set-up. The source is depicted on the left-hand side, and the measurement section on the right-hand side. The desired quantum state is generated via type-I spontaneous parametric downconversion using two orthogonally oriented bismuth triborate (BiBO) crystals. The pump beam for the downconversion process is a mixture of a decohered state that is obtained from the long arm of the interferometer and a diagonally polarized state from the short arm. The measurement section allows for tomography to be carried out when the motorized mirrors are removed and the photons traverse the beam displacer (BD) interferometers. Alice and Bob perform projective measurements when the quarter-wave plates (QWPs) of the tomography stages are removed. Alternatively, they can ask Charlie and Debbie for their respective measurement outcomes by sliding in the motorized mirrors, using the fact that the projective measurements of their friends correspond to the beam paths inside the interferometers. NPBS, non-polarizing beamsplitter; KTP, potassium titanyl phosphate; HWP, half-wave plate; APD, avalanche photodiode; PC, polarization control; PBS, polarizing beamsplitter. Figure reproduced with permission from ref. [40], SPIE.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41567-020-0990-x.

Received: 15 September 2019; Accepted: 1 July 2020;
Published online: 17 August 2020

References

1. Wigner, E. P. in *The Scientist Speculates* (ed. Good, I. J.) 284–302 (Heinemann, 1961).
2. Schlosshauer, M. Decoherence, the measurement problem, and interpretations of quantum mechanics. *Rev. Mod. Phys.* **76**, 1267–1305 (2005).
3. Leggett, A. J. The quantum measurement problem. *Science* **307**, 871–872 (2005).
4. Everett, H. 'Relative state' formulation of quantum mechanics. *Rev. Mod. Phys.* **29**, 454–462 (1957).
5. Rovelli, C. Relational quantum mechanics. *Int. J. Theor. Phys.* **35**, 1637–1678 (1996).
6. Fuchs, C. A. & Schack, R. Quantum-Bayesian coherence. *Rev. Mod. Phys.* **85**, 1693–1715 (2013).
7. Mermin, N. D. Physics: QBism puts the scientist back into science. *Nature* **507**, 421–423 (2014).
8. Bohm, D. A suggested interpretation of the quantum theory in terms of 'hidden' variables. I. *Phys. Rev.* **85**, 166–179 (1952).
9. Bohm, D. A suggested interpretation of the quantum theory in terms of 'hidden' variables. II. *Phys. Rev.* **85**, 180–193 (1952).
10. Price, H. Toy models for retrocausality. *Stud. Hist. Philos. Sci. B Mod. Phys.* **39**, 752–761 (2008).
11. 't Hooft, G. The free-will postulate in quantum mechanics. Preprint at https://arxiv.org/abs/quant-ph/0701097 (2007).
12. Bassi, A. & Ghirardi, G. Dynamical reduction models. *Phys. Rep.* **379**, 257–426 (2003).
13. Penrose, R. On gravity's role in quantum state reduction. *Gen. Relat. Gravit.* **28**, 581–600 (1996).
14. Brukner, Č. A no-go theorem for observer-independent facts. *Entropy* **20**, 350 (2018).
15. Brukner, Č. in *Quantum [Un]Speakables II: Half a Century of Bell's Theorem* (eds Bertlmann, R. & Zeilinger, A.) 95–117 (Springer, 2017).
16. Frauchiger, D. & Renner, R. Quantum theory cannot consistently describe the use of itself. *Nat. Commun.* **9**, 3711 (2018).
17. Proietti, M. et al. Experimental test of local observer independence. *Sci. Adv.* **5**, eaaw9832 (2019).
18. Baumann, V. & Wolf, S. On formalisms and interpretations. *Quantum* **2**, 99 (2018).
19. Healey, R. Quantum theory and the limits of objectivity. *Found. Phys.* **48**, 1568–1589 (2018).
20. Baumann, V., Del Santo, F. & Brukner, Č. Comment on Healey's 'Quantum theory and the limits of objectivity'. *Found. Phys.* **49**, 741–749 (2019).
21. Shimony, A. in *Foundations of Quantum Mechanics in the Light of New Technology* (ed. Kamefuchi, S.) 225–230 (Physical Society of Japan, 1984).
22. Kochen, S. & Specker, E. P. The problem of hidden variables in quantum mechanics. *J. Math. Mech.* **17**, 59–87 (1967).
23. Liang, Y.-C., Spekkens, R. W. & Wiseman, H. M. Specker's parable of the overprotective seer: a road to contextuality, nonlocality and complementarity. *Phys. Rep.* **506**, 1–39 (2011).
24. Wiseman, H. M. & Cavalcanti, E. G. in *Quantum [Un]Speakables II: Half a Century of Bell's Theorem* (eds Bertlmann, R. & Zeilinger, A.) 119–142 (Springer, 2017).
25. Cavalcanti, E. G. Classical causal models for Bell and Kochen–Specker inequality violations require fine-tuning. *Phys. Rev. X* **8**, 021018 (2018).
26. Healey, R. Reply to a comment on 'Quantum theory and the limits of objectivity'. *Found. Phys.* **49**, 816–819 (2019).
27. Peres, A. Unperformed experiments have no results. *Am. J. Phys.* **46**, 745–747 (1978).
28. Bell, J. S. On the Einstein Podolsky Rosen paradox. *Physics* **1**, 195–200 (1964).
29. Brunner, N., Cavalcanti, D., Pironio, S., Scarani, V. & Wehner, S. Bell nonlocality. *Rev. Mod. Phys.* **86**, 419–478 (2014).
30. Giustina, M. et al. Significant-loophole-free test of Bell's theorem with entangled photons. *Phys. Rev. Lett.* **115**, 250401 (2015).
31. Hensen, B. et al. Loophole-free Bell inequality violation using electron spins separated by 1.3 kilometres. *Nature* **526**, 682–686 (2015).
32. Shalm, L. K. et al. Strong loophole-free test of local realism. *Phys. Rev. Lett.* **115**, 250402 (2015).

33. Woodhead, E. *Imperfections and Self Testing in Prepare-and-Measure Quantum Key Distribution*. PhD thesis, Univ. libre de Bruxelles (2014).
34. Colbeck, R. *Quantum and Relativistic Protocols for Secure Multi-Party Computation*. PhD thesis, Univ. of Cambridge (2006).
35. Pironio, S. et al. Random numbers certified by Bell's theorem. *Nature* **464**, 1021–1024 (2010).
36. Acín, A. & Masanes, L. Certified randomness in quantum physics. *Nature* **540**, 213–219 (2016).
37. Collins, D. & Gisin, N. A relevant two qubit Bell inequality inequivalent to the CHSH inequality. *J. Phys. A* **37**, 1775–1787 (2004).
38. Clauser, J. F., Horne, M. A., Shimony, A. & Holt, R. A. Proposed experiment to test local hidden-variable theories. *Phys. Rev. Lett.* **23**, 880–884 (1969).
39. Barrett, J. et al. Nonlocal correlations as an information-theoretic resource. *Phys. Rev. A* **71**, 022101 (2005).
40. Bong, K.-W. et al. Testing the reality of Wigner's friend's experience. *Proc. SPIE* **11200**, 112001C (2019).

Methods

Quantum-mechanical description of the EWFS. We consider two superobservers, Alice and Bob, and their respective friends, Charlie and Debbie. Charlie and Debbie are in possession of systems S_A and S_B respectively, with associated Hilbert spaces $\mathcal{H}_{S_A}$ and $\mathcal{H}_{S_B}$, and initially prepared in a (possibly entangled) state $\rho_{S_AS_B}$. For simplicity, we suppose these systems are spin-1/2 particles, and that they perform a measurement of the z-spin of their particles. We denote everything in Charlie's lab except S_A as system F_A, with Hilbert space $\mathcal{H}_{F_A}$, and F_B, $\mathcal{H}_{F_B}$ for Debbie's lab.

According to Alice, Charlie's measurement of S_A in the basis $\{|-1\rangle_{S_A}, |+1\rangle_{S_A}\}$ is described by a unitary U_{Z_A} acting on $\mathcal{H}_{F_A}\otimes\mathcal{H}_{S_A}$. Alice's $x=1$ measurement (corresponding to opening the box and asking Charlie what he saw) can be described by a POVM $\{|c\rangle\langle c|_{F_A}\otimes I_{S_A}\}_c$, where $|c\rangle_{F_A}$ ($c\in\{-1,+1\}$) represents the state of Charlie after seeing outcome c and I_{S_A} is the identity operator on $\mathcal{H}_{S_A}$. The theorem makes no assumption about the form of the measurements that Alice performs for $x\in\{2,3\}$, but in our experimental realization, we consider the class of measurements that reverse the evolution U_{Z_A} that entangled F_A with S_A (Fig. 2b), followed by a measurement on S_A alone (Fig. 2c). This can be described by a POVM with elements $U_{Z_A}(I_{F_A}\otimes E_{S_A}^{a|x})U_{Z_A}^{-1}$, where I_{F_A} is the identity on $\mathcal{H}_{F_A}$ and $E_{S_A}^{a|x}$ is the positive operator associated with outcome a for measurement x that Alice performs directly on S_A. Bob's POVM elements are defined analogously. Thus, the maximum violation of the inequalities can be sought simply in measurements acting on the Hilbert spaces $\mathcal{H}_{S_A}$ and $\mathcal{H}_{S_B}$; given that Charlie and Debbie start in a known product state in the Hilbert space of $\mathcal{H}_{F_A}\otimes\mathcal{H}_{F_B}$, there is no advantage in considering arbitrary measurements on $\mathcal{H}_{F_A}\otimes\mathcal{H}_{S_A}$ and $\mathcal{H}_{F_B}\otimes\mathcal{H}_{S_B}$.

LHV correlations as a subset of LF correlations. Recall that a set of correlations has a LHV model if and only if there exists a probability distribution $P(\lambda)$ over a set of variables $\lambda\in\Lambda$ such that

$$\wp(ab|xy) = \Sigma_{\lambda\in\Lambda}P(a|x\lambda)P(b|y\lambda)P(\lambda) \tag{2}$$

for all values of the variables a, b, x, y. We now derive the general form for an LF model. From AOE and NSD, we have that

$$\wp(ab|xy) \overset{\text{AOE}}{=} \sum_{c,d} P(abcd|xy) \overset{\text{NSD}}{=} \sum_{c,d} P(ab|cdxy)P(cd) \tag{3}$$

From Locality, we can decompose the first term on the right-hand side in two ways:

$$P(ab|cdxy) = P(a|bcdxy)P(b|cdxy) \overset{\text{L}}{=} P(a|bcdxy)P(b|cdy) \tag{4}$$

or

$$P(ab|cdxy) = P(a|cdxy)P(b|acdxy) \overset{\text{L}}{=} P(a|cdx)P(b|acdxy) \tag{5}$$

Note, however, that we cannot further reduce these expressions with Locality alone, reinforcing the fact that Locality is a weaker assumption than local causality (which leads to an LHV model). However, by construction, when $x=1$ we have $a=c$, and when $y=1$, $b=d$. Then, if $x=1$, $P(a|bcdxy)=\delta_{a,c}$, and if $y=1$, $P(b|acdxy)=\delta_{b,d}$. When taking this, along with equations (4) and (5), into account, we obtain from equation (3)

$$\wp(ab|xy) = \begin{cases} \sum_{c,d}\delta_{a,c}P(b|cdy)P(cd) & \text{if } x=1 \\ \sum_{c,d}\delta_{b,d}P(a|cdx)P(cd) & \text{if } y=1 \\ \sum_{c,d}P_{\text{NS}}(ab|cdxy)P(cd) & \text{if } x\neq 1,\ y\neq 1 \end{cases} \tag{6}$$

where $P_{\text{NS}}(ab|cdxy)$ denotes some joint probability distribution that satisfies the condition of Locality. For any fixed values of c and d, it is easy to see that the set of $P_{\text{NS}}(ab|cdxy)$ is simply the no-signalling polytope with one less measurement setting for both Alice and Bob[39] (thus the NS subscript). In general, because of the additional structure given by the first two lines of equation (6), the set of LF correlations only forms a subset of the no-signalling polytope.

To see that LHV correlations are also LF correlations, we first recall from ref. [29] that correlations of the form of equation (2) can always be decomposed in terms of the extreme points of the set of such correlations. To this end, it is expedient to write the hidden variable as $\lambda=(\lambda_1^A,\lambda_1^B,\lambda_2^A,\ldots\lambda_N^B)$, with λ_x^A and λ_y^B parameterizing all possible local deterministic strategies, that is

$$P(a|x\lambda)=\delta_{a,\lambda_x^A},\quad P(b|y\lambda)=\delta_{b,\lambda_y^B} \tag{7}$$

We may now rewrite equation (2) as

$$\wp(ab|xy) = \sum_{\lambda}\delta_{a,\lambda_x^A}\ \delta_{b,\lambda_y^B}P(\lambda) \tag{8}$$

This is now readily cast in the form of equation (6) if we set $\lambda_1^A=c$ and $\lambda_1^B=d$. For example, if $x=1$, we get

$$\begin{aligned}\wp(ab|x=1,y) &= \textstyle\sum_{c,d,\lambda_2^A,\ldots,\lambda_N^B}\delta_{a,c}\ \delta_{b,\lambda_y^B}P(cd\lambda_2^A\ldots\lambda_N^B)\\ &= \textstyle\sum_{c,d,\lambda_y^B}\delta_{a,c}\ \delta_{b,\lambda_y^B}P(cd\lambda_y^B)\\ &= \textstyle\sum_{c,d}\delta_{a,c}\left[\sum_{\lambda_y^B}\delta_{b,\lambda_y^B}\ P(\lambda_y^B|cd)\right]P(cd)\\ &= \textstyle\sum_{c,d}\delta_{a,c}\ P(b|cdy)P(cd)\end{aligned} \tag{9}$$

which is clearly of the form given in the first line of equation (6). The proof for the $y=1$ case is completely analogous.

Similarly, for the case where $x\neq 1$, $y\neq 1$, we can again make use of $\lambda_1^A=c$, $\lambda_1^B=d$ and equation (8) to arrive at

$$\begin{aligned}\wp(ab|xy) &= \textstyle\sum_{c,d,\lambda_2^A,\ldots,\lambda_N^B}\delta_{a,\lambda_x^A}\ \delta_{b,\lambda_y^B}P(cd\lambda_2^A\ldots\lambda_N^B)\\ &= \textstyle\sum_{c,d,\lambda_x^A\lambda_y^B}\delta_{a,\lambda_x^A}\ \delta_{b,\lambda_y^B}P(cd\lambda_x^A\lambda_y^B)\\ &= \textstyle\sum_{c,d}\left[\sum_{\lambda_x^A,\lambda_y^B}\delta_{a,\lambda_x^A}\ \delta_{b,\lambda_y^B}\ P(\lambda_x^A\lambda_y^B|cd)\right]P(cd)\\ &= \textstyle\sum_{c,d}\ P(ab|cdxy)P(cd)\end{aligned} \tag{10}$$

From the second last line of equation (10) and the fact that a (b) is entirely decided by λ_x^A (λ_y^B), we see that $P(ab|cdxy)$ in the last expression satisfies the condition of locality (that is, $\sum_a P(ab|cdxy)$ does not depend on y while $\sum_b P(ab|cdxy)$ does not depend on x). Thus, starting from LHV correlations for $x\neq 1$, $y\neq 1$, we recover the last line of equation (6).

Hence any correlation that satisfies equation (2) will also satisfy equation (6). Yet, the opposite is not necessarily true. Therefore, LHV correlations are a subset of LF correlations.

Characterization of LF correlations. Consider a general scenario with N measurement settings per party, with O outcomes each. Note that we can always rewrite equation (6) in the form

$$\wp(ab|xy) = \begin{cases} \sum_\lambda \delta_{a,c(\lambda)}P_{\text{Ext}}^{(j(\lambda))}(b|y)P(\lambda) & \text{if } x=1 \\ \sum_\lambda \delta_{b,d(\lambda)}P_{\text{Ext}}^{(j(\lambda))}(a|x)P(\lambda) & \text{if } y=1 \\ \sum_\lambda P_{\text{Ext}}^{(j(\lambda))}(ab|xy)P(\lambda) & \text{otherwise} \end{cases} \tag{11}$$

where λ is a variable that determines the values of $c(\lambda)$, $d(\lambda)$ and that of a variable $j(\lambda)$ that labels the (finitely many) extreme points of the no-signalling polytope with $N-1$ inputs and O outputs per party, and $P_{\text{Ext}}^{(j)}(a|x)=\sum_b P_{\text{Ext}}^{(j)}(ab|xy)$ and $P_{\text{Ext}}^{(j)}(b|y)=\sum_a P_{\text{Ext}}^{(j)}(ab|xy)$ are the marginal distributions of these extremal boxes.

It is easy to see from the above that this set of correlations is convex. That is, for any two points $\wp_1(ab|xy)$and $\wp_2(ab|xy)$, both satisfying the LF conditions, any convex combination $\wp'(ab|xy)=\alpha\wp_1(ab|xy)+(1-\alpha)\wp_2(ab|xy)$, with $0<\alpha<1$, also satisfies those conditions. The set of LF correlations is therefore a polytope.

For the two-measurement-setting case ($N=2$), the $P_{\text{Ext}}^{(j)}(ab|xy)$ now refer only to the case $x=y=2$, and the extreme points are now simply deterministic functions for a, b. Thus, we recover an LHV model for any value of O, yielding the same inequalities Brukner derived for $N=O=2$.

Next, we consider the LF polytope for the $N=3$, $O=2$ scenario. Without loss of generality, we label the outcomes as a, $b\in\{+1,-1\}$. From equation (11), the set of LF correlations $\overrightarrow{\wp}=\{\wp(ab|xy)\}_{a,b=\pm1;x,y=1,2,3}$ is the convex hull of the extreme points $\{\overrightarrow{P}^{(\lambda)}(ab|xy)\}_\lambda$ defined by

$$P^{(\lambda)}(ab|xy) = \begin{cases} \delta_{a,c(\lambda)}\delta_{b,d(\lambda)} : x=y=1 \\ \delta_{a,c(\lambda)}P_{\text{Ext}}^{(j(\lambda))}(b|y) : x=1,\ y\neq 1 \\ P_{\text{Ext}}^{(j(\lambda))}(a|x)\delta_{b,d(\lambda)} : x\neq 1,\ y=1 \\ P_{\text{Ext}}^{(j(\lambda))}(ab|xy) : x\neq 1,\ y\neq 1 \end{cases} \tag{12}$$

Because there are four combinations of (c,d) corresponding to 2^2 local deterministic strategies for the first inputs, and 24 extreme points for the aforementioned no-signalling polytope[39], we thus end up with 96 points in this set.

By writing the components of these points in a text file and feeding the latter into the freely available software PANDA—which allows one to transform between the two representations of a polytope using the parallel adjacency decomposition algorithm[41]—we obtain the complete set of 932 LF facets for this scenario. Many of these inequalities can be transformed from one to another under a relabelling of parties (Alice $\leftrightarrow$ Bob), inputs ($x=2\leftrightarrow x=3$ and/or $y=2\leftrightarrow y=3$) and/or outputs ($a=+1\leftrightarrow a=-1$ and/or $b=+1\leftrightarrow b=-1$). With the exception of the settings for $x=1$ and $y=1$, the rest of these labellings are arbitrary. Taking advantage of this arbitrariness, we may group the obtained facets into the following nine inequivalent classes (written in terms of correlators, where A_i is a random variable representing the measurement result for $x=i$ and taking values $\{-1,+1\}$; similarly for B_j):

1. Genuine LF facet 1 (appearing 256 times among the 932 facets):

$$\begin{aligned}&-\langle A_1\rangle-\langle A_2\rangle-\langle B_1\rangle-\langle B_2\rangle\\ &-\langle A_1B_1\rangle-2\langle A_1B_2\rangle-2\langle A_2B_1\rangle+2\langle A_2B_2\rangle\\ &-\langle A_2B_3\rangle-\langle A_3B_2\rangle-\langle A_3B_3\rangle-6\overset{\text{LF}}{\leq}0\end{aligned} \tag{13}$$

2. Genuine LF facet 2 (appearing 256 times):

$$\begin{aligned}&-\langle A_1\rangle-\langle A_2\rangle-\langle A_3\rangle-\langle B_1\rangle\\ &-\langle A_1B_1\rangle-\langle A_2B_1\rangle-\langle A_3B_1\rangle-2\langle A_1B_2\rangle\\ &+\langle A_2B_2\rangle+\langle A_3B_2\rangle-\langle A_2B_3\rangle+\langle A_3B_3\rangle-5\overset{\text{LF}}{\leq}0\end{aligned} \tag{14}$$

3. Bell I_{3322}[37] with marginals over inputs 1 and 2 (appearing 256 times):

$$\begin{aligned}&-\langle A_1\rangle+\langle A_2\rangle+\langle B_1\rangle-\langle B_2\rangle\\&+\langle A_1B_1\rangle-\langle A_1B_2\rangle-\langle A_1B_3\rangle-\langle A_2B_1\rangle\\&+\langle A_2B_2\rangle-\langle A_2B_3\rangle-\langle A_3B_1\rangle-\langle A_3B_2\rangle-4\overset{\mathrm{LF}}{\leq}0\end{aligned}\tag{15}$$

4. Bell I_{3322} with marginals over inputs 2 and 3 (appearing 64 times):

$$\begin{aligned}&-\langle A_2\rangle-\langle A_3\rangle-\langle B_2\rangle-\langle B_3\rangle\\&-\langle A_1B_2\rangle+\langle A_1B_3\rangle-\langle A_2B_1\rangle-\langle A_2B_2\rangle\\&-\langle A_2B_3\rangle+\langle A_3B_1\rangle-\langle A_3B_2\rangle-\langle A_3B_3\rangle-4\overset{\mathrm{LF}}{\leq}0\end{aligned}\tag{16}$$

5. 'Brukner inequality': Bell-CHSH for inputs 1 and 2 of Alice and inputs 1 and 3 of Bob (appearing 32 times):

$$\langle A_1B_1\rangle-\langle A_1B_3\rangle-\langle A_2B_1\rangle-\langle A_2B_3\rangle-2\overset{\mathrm{LF}}{\leq}0\tag{17}$$

6. 'Semi-Brukner' inequality: Bell-CHSH for inputs 2 and 3 of Alice and inputs 1 and 2 of Bob (appearing 32 times):

$$-\langle A_1B_2\rangle+\langle A_1B_3\rangle-\langle A_3B_2\rangle-\langle A_3B_3\rangle-2\overset{\mathrm{LF}}{\leq}0\tag{18}$$

7. Positivity for input 1 of Alice and input 1 of Bob (appearing four times):

$$1+\langle A_1\rangle+\langle B_1\rangle+\langle A_1B_1\rangle\geq 0\tag{19}$$

8. Positivity for input 1 of Alice and input 2 of Bob (appearing 16 times):

$$1+\langle A_1\rangle+\langle B_2\rangle+\langle A_1B_2\rangle\geq 0\tag{20}$$

9. Positivity for input 2 of Alice and input 2 of Bob (appearing 16 times):

$$1+\langle A_2\rangle+\langle B_2\rangle+\langle A_2B_2\rangle\geq 0\tag{21}$$

Note that some Bell facets for this scenario are not facets of LF and thus do not appear in the list above, for example, the Bell-CHSH inequalities that do not include any input 1 for either party:

$$\langle A_2B_2\rangle-\langle A_2B_3\rangle-\langle A_3B_2\rangle-\langle A_3B_3\rangle-2\overset{\mathrm{LHV}}{\leq}0\tag{22}$$

Inequalities and measurements considered in the experiment. For each category, the inequalities we considered in our experiment were genuine LF (equation (13)), I_{3322} (equation (15)), Brukner (equation (17)), semi-Brukner (equation (18)) and Bell non-LF (equation (22)).

Here we use $A_x\in\{+1,-1\}$ as the random variable for Alice's outcome a when she chooses setting x, and similarly B_y. That is, the expectation values are calculated from the empirical probabilities $\wp(ab|xy)$.

We restrict ourselves to projective measurements in the X–Y plane of the Bloch sphere (with states $|H\rangle$ and $|V\rangle$ on the z axis). In particular, Alice's measurement results are represented by operators of the form $A_x=2\Pi_x^{a=1}-|H\rangle\langle H|-|V\rangle\langle V|$, with $\Pi_x^{a=1}=|\phi_x\rangle\langle\phi_x|$ being the projector onto the state

$$|\phi_x\rangle=\frac{1}{\sqrt{2}}\left(|H\rangle+\mathrm{e}^{\mathrm{i}\phi_x}|V\rangle\right)\tag{23}$$

Bob's corresponding operators are chosen to be $B_y=2\Pi_y^{b=1}-|H\rangle\langle H|-|V\rangle\langle V|$, with $\Pi_y^{b=1}=|\beta_y\rangle\langle\beta_y|$ being the projector onto

$$|\beta_y\rangle=\frac{1}{\sqrt{2}}\left(|H\rangle+\mathrm{e}^{\mathrm{i}(\beta-\phi_y)}|V\rangle\right)\tag{24}$$

For each value of the tetrad $(\phi_1,\phi_2,\phi_3,\beta)$, and for each category in Table 1, we find the smallest value of μ for which one of the inequalities in that category is violated. We then pick a tetrad that makes the gap between these values of μ conveniently large. The values we choose are $\phi_1=168°$, $\phi_2=0°$, $\phi_3=118°$ and $\beta=175°$. For the inequality in each category that is violated first, we display the values of the left-hand side as a function of μ in Fig. 4.

Spontaneous parametric downconversion source. The source is made up of an imbalanced pump-beam interferometer (one arm of the interferometer is longer than the other) and two orthogonally oriented (sandwiched) BiBO crystals[42], which are pumped by a 404-nm continuous-wave laser diode to produce spontaneous parametric downconversion. The relative pump power in the interferometer arms determines the μ parameter of the state and is controlled by the HWP after the laser. When all the pump power is in the short arm, the first term of the quantum state, the singlet state, is generated (after a local polarization rotation in the fibre). Conversely, when all the pump power is in the long arm, only the second term, a mixed state, is generated. The beams in both arms are recombined in the NPBS to pump the sandwiched crystal, generating the desired quantum state.

The polarization in the short arm is rotated to diagonal by a HWP and an additional birefringent element is used to pre-compensate the temporal walk-off in the downconversion. The polarization in the long arm is also rotated to diagonal by a HWP and a birefringent crystal decoheres the horizontal and vertical polarization components completely, which is necessary to generate the mixed part of the state.

Quantum state tomography. To allow tomography, the motorized mirrors are moved out of the beam paths and the measurements are carried out using the last QWP, HWP and PBS on each side. As part of the tomographic state reconstruction, the known unitary transformations of the first QWP and the BD interferometer are accounted for, such that the quantum state straight after the fibre is obtained. Typically, ~22,000 coincidences are collected per tomography. The implemented μ values are estimated by comparing the reconstructed states with the set of target states ρ_μ, and finding the μ values that maximize the fidelity (for details, see Supplementary Section D).

Experimental implementation of measurements. The measurements of the EWFS are realized in the following way. When measurement setting 1 is chosen, the motorized mirror is inserted to reveal the photon path within the interferometer, that is after BD1, and this corresponds to Alice asking Charlie his measurement outcome (or Bob asking Debbie on the other side). This comprises the first of the possible POVMs, illustrated in Fig. 2a. When measurement settings 2 or 3 are chosen, Alice (Bob) first reverses Charlie's (Debbie's) measurement (Fig. 2b) by removing the mirror between the two BDs and thereby closing the interferometer, and then proceeds to measure the polarization after the interferometer with the QWP after BD2 removed (Fig. 2c). This two-step procedure corresponds to Alice (Bob) implementing one of her (his) other two POVMs, depending on which one of two settings of the last HWP is used. Single photons are detected with APDs and coincidences are recorded with counting modules. The overall observed rate of counts in the apparatus is ~550 coincidences and 21,000 singles per second.

To obtain the expectation values required for the inequalities being tested at each μ value, we performed the nine sets of measurements that arise from combining the three independent measurement settings on Alice's and Bob's sides. The typical number of counts per measurement set is 91,000 coincidences.

Data availability

Data that support the plots within this paper and other findings of this study are available from the corresponding authors upon reasonable request. Source data are provided with this paper.

Code availability

The numerical codes used to determine the inequalities and to choose the measurement settings are available from the corresponding authors upon reasonable request.

References

41. Lörwald, S. & Reinelt, G. PANDA: a software for polyhedral transformations. *EURO J. Comput. Optim.* **3**, 297–308 (2015).
42. Altepeter, J. B., Jeffrey, E. R. & Kwiat, P. G. Phase-compensated ultra-bright source of entangled photons. *Opt. Express* **13**, 8951–8959 (2005).

Acknowledgements

This work was supported by the Australian Research Council (ARC) Centre of Excellence CE170100012, the Ministry of Science and Technology, Taiwan (grant nos. 107-2112-M-006-005-MY2 and 107-2627-E-006-001), ARC Future Fellowship FT180100317 and grant no. FQXi-RFP-1807 from the Foundational Questions Institute and Fetzer Franklin Fund, a donor advised fund of Silicon Valley Community Foundation. A.U.-A., K.-W.B. and F.G. acknowledge financial support through Australian Government Research Training Program Scholarships and N.T. acknowledges support by the Griffith University Postdoctoral Fellowship Scheme. We gratefully acknowledge A. Acín for bringing ref. [33] to our attention, and thank S. Slussarenko for useful discussions. Avatars in Figs. 1 and 2 are adapted from Eucalyp Studio, available under a Creative Commons licence (Attribution 3.0 Unported), https://creativecommons.org/licenses/by/3.0/, at https://www.iconfinder.com/iconsets/avatar-55.

Author contributions

A.U.-A., E.G.C., Y.-C.L. and H.M.W. performed the theory work. K.-W.B., N.T., H.M.W. and G.J.P. designed the experiment, which was realized by K.-W.B., N.T., F.G. and G.J.P. All authors contributed to the preparation of the manuscript and N.T. and E.G.C. took responsibility for its final form.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information is available for this paper at https://doi.org/10.1038/s41567-020-0990-x.

Correspondence and requests for materials should be addressed to N.T. or E.G.C.

Reprints and permissions information is available at www.nature.com/reprints.